

# Validation of the multidimensional bystander responses to racist hate speech scale and its association with empathy and moral disengagement among adolescents

Sebastian Wachs<sup>1,2</sup>  | Ludwig Bilz<sup>3</sup>  | Alexander Wettstein<sup>4</sup> | Dorothy L. Espelage<sup>5</sup>

<sup>1</sup>Department of Educational Sciences, University of Potsdam, Potsdam, Germany

<sup>2</sup>Institute of Education, University of Münster, Münster, Germany

<sup>3</sup>Department of Health Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany

<sup>4</sup>Institute for Research, Development and Evaluation, Bern University of Teacher Education, Bern, Switzerland

<sup>5</sup>School of Education, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

## Correspondence

Sebastian Wachs, Department of Educational Sciences, University of Potsdam, Karl-Liebknecht, Str. 24–25, 14476 Potsdam, Germany.  
Email: [swachs@uni-muenster.de](mailto:swachs@uni-muenster.de)

## Funding information

German Research Foundation (DFG), Grant/Award Numbers: WA4275/2-1, BI1046/9-1; University of Teacher Education Bern, Grant/Award Number: 19s000801

## Abstract

Our understanding of how bystanders respond to hate speech is limited. This may be due, in part, to the lack of available measurement tools. However, understanding adolescents' responses to hate speech is critical because this kind of research can support schools in empowering students to exhibit courageous moral behavior. Thus, the purpose of the present study was to investigate the psychometric properties of the newly developed Multidimensional Bystander Responses to Hate Speech Scale (MBRHS) and to explore demographic differences and correlates of bystander behavior in school hate speech. The sample consisted of 3225 seventh to ninth graders from Germany and Switzerland. Exploratory and confirmatory factor analyses supported a model with seven factors. We found that adolescents with immigrant background and boys showed particularly unfavorable response patterns. In addition, our study suggests that empathy is positively correlated with the factors *comforting the victim*, *seeking help at school*, and *countering hate speech* but negatively correlated with *helplessness*, *revenge*, *reinforcing*, and *ignoring*. Moral disengagement showed the opposite correlational pattern. The findings indicate that the MBRHS is a psychometrically valid and reliable measure that could aid in measuring varied responses to hate speech. In addition, this work highlights the relevance of empathy and moral engagement training in anti-hate speech prevention programs.

## KEYWORDS

adolescents, civic education, counter speech, empathy, hate speech, measurement invariance, discrimination

## 1 | INTRODUCTION

Against the background of a more visibly diverse society, racial prejudices have flourished in Germany over the last few years (Bundesministerium des Innern und Heimat BMI, 2023). Strongly related to this is hate speech, which can be understood as a

behavioral component of prejudice. Hate speech can be defined as any derogatory expression (e.g., words, posts, images, videos) expressed to cause harm to people (directly or vicariously) based on actual or assigned group characteristics (e.g., ethnicity, sexual orientation, religious affiliation; Kansok-Dusche et al., 2022). The most common targets of hate speech in German schools are people

[Correction added on 6 October 2023, after first online publication: the affiliation 2 and email address for author "Sebastian Wachs" has been corrected.]

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Aggressive Behavior* published by Wiley Periodicals LLC.

from ethnic minorities and with immigrant background (Castellanos, Wettstein, Wachs, Kansok-Dusche, et al., 2023). Therefore, we will focus on racist hate speech in this study. Research on hate speech is needed because there is some evidence that hate speech victimization can impair adolescents' well-being (Krause et al., 2021) and the classroom climate (Ballaschk et al., 2021). In bullying research, it is essential to consider bystanders when attempting to understand the occurrence of bullying in classrooms and to ensure the success of bullying interventions (Nocentini et al., 2013). In hate speech research, however, there is no validated scale to capture the variety of possible bystander responses. Thus, our understanding of the role of the bystander is limited. To narrow these gaps in the literature, this study examines the psychometric properties of a newly developed instrument to measure bystanders' responses to racist hate speech. It also explores the demographic differences in bystanders' responses and investigates the scale's associations with empathy and moral disengagement. The findings might help schools create an inclusive and supportive environment and bring the research field forward by delivering tools to measure adolescent hate speech.

### 1.1 | Bystander responses to hate speech

In one study among German adolescents, around 65% of the participants reported witnessing hate speech in schools at least once within 12 months (Castellanos, Wettstein, Wachs, Kansok-Dusche, et al., 2023). And yet, hate speech research has focused solely on the reactions of adolescents who became targets of hate speech. For example, in a quantitative study with German adolescents, victims reported that the most frequent coping strategy to deal with online hate speech was technical coping, followed by assertiveness, seeking close support, helplessness/self-blame, retaliation, and seeking distal advice (Wachs et al., 2020). In a qualitative study, adolescents reported several responses to hate speech victimization in schools, such as seeking peer support, countering hate speech, avoiding people who use hate speech, and ignoring the incident (Krause et al., 2021).

In bullying research, bystander responses to bullying incidents have been investigated thoroughly. In their seminal work, Salmivalli et al. (1996) described various bystander responses to bullying, including assistants who actively assist the bullies and reinforcers who remain present and show their support for the bullying through laughter or cheering. Defenders take action to protect the victimized individuals or provide support for them, while outsiders remain passive and do not explicitly indicate approval or disapproval of the situation. Other scholars distinguish between three common bystander responses (e.g., Álvarez-García et al., 2021; Thornberg et al., 2022): (1) *defender behaviors*, which include, for example, comforting the targeted person, seeking support from peers or adults, and confronting the perpetrator; (2) *pro-bully behaviors*, which include assisting the perpetrator, or reinforce by laughing about it; and (3) *passive behaviors*, which include, for example, ignoring the incident or feeling helpless about one's ability to do something about it. Regarding frequencies, a study on bias-based bullying found that

the most common response of adolescents who witness race-based bullying toward immigrants was talking to the victim, followed by saying something to the aggressor, seeking help from peers and adults, and not getting involved (Gönültaş & Mulvey, 2021).

### 1.2 | Differences in bystander responses to hate speech as a function of demographic variables

Research from related fields (e.g., online hate speech and bullying) have shown contradicting findings regarding differences in adolescents' responses to online hate speech victimization as a function of age, gender, and immigrant background. In terms of age, a study on how Spanish adolescents cope with online hate speech showed that younger adolescents were more likely to use social support than older adolescents, and young adolescents were more likely to feel helpless when dealing with online hate speech than older adolescents (Gámez Guadix et al., 2020). In bullying research, studies have revealed either no association between age and any typical bystander responses (Mazzone et al., 2016; Shen et al., 2022), a positive relationship between age and prosocial bystander responses, or a negative link between age and antisocial bystander behavior (Schultze-Krumbholz et al., 2018; Yang & Kim, 2017) and passive bystander behavior (Salmivalli & Voeten, 2004; Schultze-Krumbholz et al., 2018).

Findings on gender differences in bullying bystanders are similarly mixed. While some research reported no gender differences in the three categories of bystander responses outlined above (Shen et al., 2022), other research has suggested that, respectively, either boys (Schultze-Krumbholz et al., 2018) or girls (Mazzone et al., 2016; Thornberg et al., 2022) were more likely to show defender bystander behavior. Regarding pro-bully bystander behavior, research has often suggested that boys are more likely than girls to engage in this kind of behavior (Schultze-Krumbholz et al., 2018; Yang & Kim, 2017), but there is also research that has not been able to find gender differences in bystander responses (e.g., Shen et al., 2022).

Evidence shows that bystander responses are often influenced by individuals' social group memberships and social identities. For example, one study showed that adolescents without immigrant background were less likely than adolescents with immigrant background to intervene in bias-based bullying when the target was a peer with an immigrant background (Gönültaş & Mulvey, 2021). Another study revealed that bystanders without immigrant background were more likely to intervene prosocially when the excluded peer had no immigrant background, but bystanders with immigrant background showed defender bystander behavior regardless of the excluded peer's immigrant background (Palmer et al., 2022).

### 1.3 | Empathy, moral disengagement, and bystander responses to hate speech

Empathy and moral withdrawal have often been highlighted in the context of prosocial and antisocial behavior. Empathy can be defined as

the ability to understand and feel the cognitive and affective experiences of others (Batson, 2009). Empathy is crucial in prosocial and antisocial behavior, intolerance, and prejudice (Boag & Carnelley, 2016; Pettigrew & Tropp, 2008). In hate speech research, higher levels of empathy are negatively associated with online and offline hate speech perpetration and the acceptance of online hate speech (Castellanos, Wettstein, Wachs, & Bilz, 2023; Celuch et al., 2022; Wachs, Bilz, et al., 2022). In addition to this, people with high levels of empathy are more likely to perceive hate speech as harmful (Cowan & Khatchadourian, 2003). While there is a lack of studies investigating bystander responses to hate speech in school settings, several studies have examined the links between empathy and bystander responses to (bias-based) traditional bullying and cyberbullying. These studies found unequivocally that empathy was positively related to defender reactions in favor of the targeted person (Álvarez-García et al., 2021; Gönültaş & Mulvey, 2022; Shen et al., 2022; Yang & Kim, 2017) and negatively correlated with pro-bully or passive bystander behavior (Álvarez-García et al., 2021; Gönültaş & Mulvey, 2022; Schultze-Krumbholz et al., 2018; Shen et al., 2022; Yang & Kim, 2017). Hence, it can be assumed that more empathic adolescents are better able to show compassion to the targeted person and, thus, are more willing to show defender bystander responses and less likely to display antisocial or passive bystander behavior in response to hate speech.

Moral disengagement refers to the selective or total deactivation of several socio-cognitive mechanisms that serve as a strategy to reduce guilt and self-censoring emotions, thus allowing people to justify their harmful behavior toward others (Bandura et al., 1996). Mechanisms that enable moral disengagement include justification through attributing blame to the target, dehumanizing the target, distorting the negative consequences for the target, minimizing agency, and cognitive restructuring (Bandura et al., 1996). Initial research has shown that people with high levels of moral disengagement are more likely to perpetrate online hate speech and perceive this behavior as acceptable (D'Errico & Paciello, 2018; Wachs et al., 2022). Like empathy, moral disengagement is related to various bystander responses to traditional bullying and cyberbullying. Specifically, this research showed that moral disengagement is negatively related to defending in favor of the victim and positively associated with pro-bully or passive bystander behavior (Killer et al., 2019; Raboteg-Šaric, 2019; Shen et al., 2022; Thornberg et al., 2022). Hate speech bystanders with high moral disengagement might, thus, be less likely to consider the suffering that hate speech inflicts on others, more likely not to feel burdened by self-sanctions, and less concerned about moral reasoning because of moral disengagement. Therefore, it can be assumed that adolescents who feel morally disengaged might be less likely to show prosocial bystander responses but more likely to exhibit antisocial and passive bystander behavior than adolescents who do not.

## 1.4 | The present study

Bearing in mind the scant availability of scales for measuring bystander responses to hate speech that yields reliable and valid

scores, the present study's first aim was to investigate the validity and reliability of a new multidimensional scale to measure bystander responses to hate speech among students. The second aim was to examine measurement invariance and investigate latent mean differences as a function of grade, gender, and immigrant background. Given the contradicting findings of previous research in related research areas, exploring demographic differences is necessarily exploratory. The third aim was to examine the relationship between empathy, moral disengagement, and the MBRHS.

## 2 | METHODS

### 2.1 | Participants

The present sample is based on 3225 adolescents (approximately between 12 and 15 years old) from Germany ( $n = 1841$ ; 57.1%) and Switzerland ( $n = 1384$ ; 42.9%). Participants were in grades 7–9 (seventh grade: 33.2%,  $n = 1070$ ; eighth grade: 35.6%,  $n = 1147$ ; ninth grade: 31.3%,  $n = 1008$ ). In terms of gender, 46.1% ( $n = 1487$ ) self-identified as boys, 51.7% ( $n = 1668$ ) as girls, 2% ( $n = 64$ ) as gender diverse, and 0.2 ( $n = 6$ ) did not indicate their gender. Regarding immigrant background, 40.3% ( $n = 1301$ ) had an immigrant background, and 59.7% ( $n = 1924$ ) did not have an immigrant background. In total, 30.8% ( $n = 994$ ) of students reported living in families of low affluence, 35.8% ( $n = 1155$ ) in families of medium affluence, and 32.4% ( $n = 1046$ ) in families of high affluence. For 0.9% ( $n = 30$ ) of all participants, socioeconomic status (SES) could not be established due to missing values.

### 2.2 | Measures

#### 2.2.1 | Bystander responses to racist hate speech

Participants were presented with a vignette that described a hate speech incident, which read as follows: "Please imagine the following situation: At your school, a student publicly makes insulting statements about people of a certain skin color or origin." Then the participants were asked: "What would you do in the situation described, or what have you done if you have experienced such a situation before?" Following this question, participants were asked to rate 21 items (see Table 2). All items could be answered on a five-point response scale ranging from "strongly disagree" to "strongly agree."

#### 2.2.2 | Empathy and moral disengagement toward hate speech

Empathy and moral disengagement toward hate speech was measured with two scales developed by Knauf et al. (2018). We adapted this instrument by referring to hate speech in the scale's

introduction: "When I see classmates being insulted or attacked by other classmates because of their skin color, origin, religion, sexual orientation, or gender...." For empathy, six items were used (e.g., *I realize how badly they are doing; it hurts me, too*). For moral disengagement, nine items were used (e.g., *I figure it's their fault; I assume they deserve no better*). All items could be rated on a five-point scale from "absolutely disagree" to "absolutely agree." McDonald's  $\omega$  indicated good reliabilities for moral disengagement 0.90 and empathy 0.91.

#### Demographic variables

Participants were asked to identify their grades and gender (boy, girl, gender diverse). The immigrant background was assessed by asking whether the participant or one of their parents was born in a country other than Germany or Switzerland. SES was measured using the Family Affluence Scale (FAS; Hartley et al., 2016). A composite FAS score calculated an individual FAS category for each participant (i.e., low, medium, and high SES). The country of origin was assigned after data cleaning.

## 2.3 | Procedure and sampling technique

### 2.3.1 | Sampling procedure

After obtaining ethical approval for the current study from the University of Potsdam Ethics Committee, data were collected between October 2020 and April 2021 via a tablet-based questionnaire in Germany and online surveys in Switzerland. In Germany, an acquisition pool of sample schools was composed of the federal state of Berlin and Brandenburg, with the type of school (e.g., grammar secondary school [*Gymnasium*] or nonacademic-track secondary school [*Realschule*]) being stratified and randomized using the probability-proportional-to-size scheme (Yates & Grundy, 1953). In Switzerland, the acquisition pool of sample schools was designed using a contrastive sampling scheme based on high/low immigrant background and rural/urban geography. From the resulting acquisition pools, 100 schools (Germany:  $n = 76$ ; Switzerland:  $n = 24$ ) were informed via phone calls and e-mails that they had been randomly selected to participate in the study. Acquisition stopped as soon as the sampling plans were fulfilled. In total, 40 schools (Germany:  $n = 18$ ; Switzerland:  $n = 22$ ) agreed to participate. The participation rate at the school level was 40% in the whole sample (Germany: 24%; Switzerland: 92%).

In the present study, seventh-, eighth-, and ninth-grade students were asked to participate in the survey. This age group was chosen because research showed that related phenomena (e.g., bullying perpetration) were most prevalent during that time among students in Germany (Fischer et al., 2020). In Germany, two randomly selected classes per grade were invited. In Switzerland, all available classes across grades 7–9 were invited. In addition, Swiss students in mixed grades were also asked to take part. In total, 264 school classes were invited to participate in the study (Germany: 106; Switzerland: 158). Of these, 236 took part in the study (Germany:  $n = 98$ ; Switzerland:

$n = 138$ ). The response rate at the classroom level was 89% for the whole sample (Germany: 92%; Switzerland: 87%).

Overall, 5836 students (Germany:  $n = 2495$ ; Switzerland:  $n = 3341$ ) were invited to participate in the current study, and 3560 students participated (Germany:  $n = 1841$ ; Switzerland:  $n = 1719$ ). The response rate at the student level was 61% for the whole sample (Germany: 74%; Switzerland: 51%). In total, 335 Swiss students from mixed classrooms in four schools were excluded from the analyses because being in classrooms with mixed grades was confounded with being Swiss. Additional information on the sampling procedure in Germany can be found in the Supporting Information Materials.

### 2.3.2 | Scale development procedure

Step 1: We conducted 55 qualitative interviews with 10th graders from seven schools in Germany, who were diverse in terms of gender and ethnicity, to investigate their experiences with and perception of hate speech. To understand adolescents' reactions when witnessing hate speech, we asked the following questions: "What did you do when you witnessed hate speech among your classmates?", "What is the appropriate way for you to address hate speech?", "What did your classmates do when hate speech occurred?" Before asking these questions, a definition of hate speech was presented to them to increase the validity of their responses. Parallel to that, we reviewed existing research instruments that measured adolescents' reactions to online hate speech, discrimination, and bullying. Step 2: Based on the findings of the qualitative study and the review of existing instruments, we developed the first version of the Multidimensional Bystander Responses to Hate Speech Scale (MBRHS), which encompassed seven subscales with 21 items. Step 3: The newly developed instrument was critically evaluated within the research group and revised accordingly. We then conducted a pre-test questionnaire with German students in grades 7–9 ( $N = 75$ ) to help determine whether students understood the items as intended. Step 4: Based on the feedback given by the students and researchers, the wording of the items was simplified, and we finalized the questionnaire and used the resulting version to survey the participants of the present study.

## 2.4 | Data analyses

### 2.4.1 | Power analysis

A priori conducted power analysis with G\*Power (Faul et al., 2007) revealed that to detect small to medium correlational effect sizes, the present study needed a sample consisting of at least 782 participants ( $\alpha = .05$ , Power = 0.80). Based on the hierarchical structure of the sample and expected nonresponse rate, the resulting minimal sample size was  $N = 1944$  students in 108 classes at 18 schools (Teerenstra et al., 2010). Accordingly, the present sample size was sufficient to investigate the research questions.

## 2.4.2 | Missing data analysis

Overall, missing data were between 1.1% ( $n = 35$ ; *ignoring*) and 1.4%, ( $n = 46$ ; *revenge*). Little's Missing Completely at Random test revealed that data were not missing at random ( $\chi^2 = 137.91$   $df = 82$ ;  $p < .001$ ). Hence, the full information maximum likelihood (FIML) estimator in *Mplus* was used to address missing data.

## 2.4.3 | Main analyses

Before conducting the analyses, the full sample ( $N = 3225$ ) was randomly split into two subsamples. Splitting the sample allows for a rigorous examination of the factor structure by validating the findings across independent subsamples. The first subsample ( $n = 1613$ ) was used for the exploratory factor analysis (EFA) to identify the best-fitting model and determine the most interpretable solution. The EFA was employed to better understand the data structure, considering that the MBRHS instrument is newly developed and its factor structure has not been examined yet. Although we had a priori assumptions about a seven-factor structure, EFA is deemed more appropriate at the initial stage of developing a new scale in case there may be unexpected but substantively meaningful factors or unexpected cross-loadings (Flora & Flake, 2017). To ensure accuracy, we employed the Maximum Likelihood estimator (MLR) with robust standard errors as the estimation method to adjust for the nonnormality and nonindependence of observations. Given the theoretical correlation between the scale items, we employed the Geomin (oblique) rotation method. Determining the number of factors retained was based on using the Kaiser criterion (eigenvalue  $> 1$ ), comparing the goodness-of-fit indices for each model obtained in the EFA, and interpreting the theoretical meaningfulness of each factor solution (Clark & Bowles, 2018; Finch, 2020).

The second subsample ( $n = 1612$ ) was utilized for conducting confirmatory factor analysis (CFA) to evaluate the model fit based on the structure derived from the EFA. This analysis aimed to assess the goodness of fit of the factor structure identified in the previous EFA. The model fit was examined in the EFA and CFA by considering the following fit indices: comparative fit index (CFI), Tucker–Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The quality of the model fit was evaluated using typical cutoff scores reflecting good and adequate fit of the data, respectively: CFI and TLI  $> 0.95$  and  $0.90$ ; RMSEA  $< 0.06$  and  $0.08$ , and SRMR  $< 0.10$  and  $0.05$  (Hu & Bentler, 1999).

Using the full sample, composite reliability (CR) and construct reliability (McDonald's  $\omega$ ) were investigated to analyze reliability. If the subscale had only two items, we used the Spearman–Brown correlation as a reliability indicator (Eisinga et al., 2013). For all reliability tests, we used 0.70 as a benchmark value for acceptable reliability (Hair et al., 2010).

Based on the full sample, we then conducted a series of multi-group confirmatory factor analyses (MGCFAs) with MLR to analyze

the measurement invariance of the MBRHS between grades, gender, and immigrant background. To compare nested models (i.e., EFA, MGCFAs), changes in goodness-of-fit indices were evaluated using Chen's (2007) recommendation, according to which decreases in  $\Delta CFI > 0.010$  and increases in  $\Delta RMSEA > 0.015$  indicate that the assumption of measurement invariance is not met. Cohen's  $d$  was used to measure effect sizes for latent factor means. Cohen's  $d$  was calculated by dividing the difference between factor means by the pooled factor standard deviation (Lakens, 2013). To evaluate the effect sizes, the typical cutoffs were used:  $d = 0.2$  (small),  $d = 0.5$  (medium), and  $d = 0.8$  (large; Cohen, 1988). Finally, latent correlations were examined to investigate the associations between empathy, moral disengagement, and the MBRHS. All analyses were conducted using *Mplus* version 8.7 (Muthén & Muthén, 2012–2021).

## 3 | RESULTS

### 3.1 | EFA, CFA, and reliability of the MBRHS

Using subsample 1, the EFA showed that seven factors had an eigenvalue  $> 1$  (4.38, 3.31, 2.42, 1.83, 1.36, 1.29, 1.09), while the eighth factor had an eigenvalue  $< 1$  (0.74), indicating that seven factors should be retained. Comparing the model fits obtained in the EFAs showed that the six- to eight-factor structure fit the data well (see Table 1). The increase in CFI and TLI of the seven-factor model compared with the six-factor model was  $> 0.010$ , and the decrease in RMSEA was  $> 0.015$ , suggesting that the seven-factor model was significantly better than the six-factor model. Comparing the seven- and eight-factor models revealed that the increase in CFI and TLI was  $< 0.010$  and the decrease in RMSEA  $< 0.015$ , indicating that the eight-factor model did not fit the data better than the seven-factor model. Therefore, we finally fixed extraction to seven factors, and based on the content of items, we named the subscales as follows *comforting the victim*, *seeking help at school*, *countering hate speech*, *helplessness*, *revenge*, *reinforcing*, and *ignoring* (Table 1).

All items across the seven factors of the MBRHS and factor loadings are presented in Table 2. No unexpected cross-loadings were observed. We then ran a CFA with the proposed seven factors with subsample 2 ( $n = 1612$ ). The seven-factor model provided an acceptable model fit,  $\chi^2 = 723.29$ ,  $df = 168$ ,  $p < .001$ , CFI = 0.955, TLI = 0.944, RMSEA = 0.042 [0.039, 0.045], SRMR = 0.041. All standardized factor loadings were significant ( $p < .001$ ) and ranged between 0.50 and 0.89 (see Table 2). The seven-factor model also provided in the full sample an acceptable model fit,  $\chi^2 = 1407.71$ ,  $df = 168$ ,  $p < .001$ , CFI = 0.956, TLI = 0.945, RMSEA = 0.048 [0.046, 0.050], SRMR = 0.041 (the factor loading can be found in Supporting Information: Table S1).

Based on the full sample, the CR was 0.84 for *comforting the victim*, 0.85 for *seeking help at school*, 0.82 for *countering hate speech*, 0.84 for *revenge*, 0.65 for *reinforcing*, 0.77 for *helplessness*, and 0.77 for *ignoring*. The McDonald's  $\omega$  was 0.84 for *comforting the victim*, 0.85 for *seeking help at school*, 0.81 for *countering hate speech*, 0.84

**TABLE 1** Fit indices for exploratory models of the MBRHS.

Model	$\chi^2$ (df)	p	RMSEA [C.I. 90%]	$\Delta$ RMSEA	CFI	$\Delta$ CFI	SRMR
1 Factor	6395.67 (189)	<.001	0.145 [0.142, 0.148]		0.416		0.135
2 Factors	4378.99 (169)	<.001	0.126 [0.123, 0.130]	-0.019	0.604	0.188	0.093
3 Factors	2989.06 (150)	<.001	0.110 [0.107, 0.114]	-0.016	0.733	0.129	0.061
4 Factors	1746.21 (132)	<.001	0.089 [0.085, 0.092]	-0.021	0.848	0.115	0.047
5 Factors	1273.62 (115)	<.001	0.080 [0.076, 0.084]	-0.009	0.891	0.043	0.036
6 Factors	415.63 (99)	<.001	0.045 [0.041, 0.050]	-0.035	0.970	0.079	0.021
7 Factors	155.22 (84)	<.001	0.023 [0.017, 0.029]	-0.022	0.993	0.023	0.011
8 Factors	109.27 (70)	<.001	0.019 [0.012, 0.026]	-0.004	0.996	0.003	0.008

Abbreviations:  $\chi^2$ , chi-square test of model fit; df, degrees of freedom; CFI, comparative fit index;  $\Delta$ CFI, change in CFI compared with the previous model above; RMSEA, root mean square error of approximation;  $\Delta$ RMSEA, change in RMSEA.

**TABLE 2** Factor loadings of the seven-factor solution of the MBRHS in EFA (subsample 1) and CFA (subsample 2).

Items		Factor							CFA
		EFA 1	2	3	4	5	6	7	
COM1	I stand by the person who is the target of hate speech.	<b>0.622</b>	-0.038	0.031	-0.038	0.076	0.019	-0.085	0.624
COM2	I comfort the person who is the target of hate speech.	<b>0.877</b>	0.001	0.012	0.007	-0.026	-0.004	0.010	0.892
COM3	I ask the person who is the target of hate speech if I could help them.	<b>0.852</b>	0.051	0.002	-0.006	-0.029	-0.026	0.027	0.891
SEEK1	I ask classmates if they can help me to do something about it.	0.165	<b>0.560</b>	0.154	0.064	-0.032	-0.042	-0.031	0.768
SEEK2	I'll report the incident to the school.	-0.018	<b>0.778</b>	0.073	-0.035	0.021	-0.020	-0.007	0.711
SEEK3	I ask my teachers if they can help me to do something about it.	0.008	<b>0.962</b>	-0.034	-0.020	0.011	-0.014	0.026	0.741
COUN1	I say to the person that such statements are hurtful.	-0.040	0.030	<b>0.838</b>	-0.042	-0.050	0.113	-0.079	0.819
COUN2	I urge the person to stop doing it.	0.016	0.005	<b>0.824</b>	0.103	-0.049	-0.017	0.007	0.846
COUN3	I try to get the person to think by asking specific questions.	0.049	-0.021	<b>0.707</b>	-0.067	0.161	-0.077	0.012	0.697
COUN4	I am saying that the person is spreading false information ("fake news").	0.082	0.042	<b>0.586</b>	0.059	0.147	-0.033	0.042	0.495
REVE1	I insult the person who carried out hate speech.	0.002	-0.110	0.067	<b>0.737</b>	0.065	-0.009	0.011	0.589
REVE2	I take revenge with others or alone on the person who uttered hate speech.	-0.005	-0.002	-0.021	<b>0.920</b>	-0.011	-0.019	0.028	0.856
REVE3	I threaten the person who uttered hate speech with violence.	0.021	0.045	-0.046	<b>0.525</b>	0.199	0.033	-0.054	0.720
REIN1	I laugh at the statement.	-0.086	-0.031	0.022	0.019	<b>0.716</b>	-0.036	0.060	0.790
REIN2	I cheer on the person making hate speech.	0.075	-0.003	-0.076	0.103	<b>0.476</b>	0.070	-0.009	0.852
REIN3	I observe the situation because I find it funny.	-0.012	0.058	0.001	0.059	<b>0.525</b>	0.048	-0.033	0.750
HELP1	I don't know what to do.	0.048	0.065	-0.008	0.029	-0.082	<b>0.728</b>	-0.060	0.781
HELP2	I don't do anything because I don't want to make it worse.	-0.008	-0.019	0.019	-0.008	0.036	<b>0.752</b>	0.080	0.500
HELP3	I don't do anything because I can't change anything about it.	-0.043	-0.056	-0.029	-0.024	0.184	<b>0.634</b>	0.087	0.598
IGN1	I just ignore it.	0.007	-0.059	-0.034	-0.050	0.047	0.048	<b>0.729</b>	0.829
IGN2	I try not to think about it.	-0.011	0.059	0.013	0.047	-0.034	0.007	<b>0.844</b>	0.709

Note: Primary loadings for each observed variable are in bold. COM, Comforting the victim; SEEK, Seeking help at school; COUN, Countering hate speech; REVE, Revenge; REIN, Reinforcing; HELP, Helplessness; IGN, Ignoring.

for *revenge*, 0.65 for *reinforcing*, and 0.77 for *helplessness*. The Spearman-Brown correlation between the two-item subscale *ignoring* was 0.76,  $p < .001$ .

### 3.2 | Frequencies of bystander responses to hate speech

The three most frequently endorsed responses (combined: *somewhat agree* and *totally agree*) in the full sample were: "I ask my teachers if they could help me to do something about it" (SEEK3; 61.2%), "I comfort the person against whom the hate speech was targeted" (COM2; 59.9%), and "I urge the person to stop doing it" (COUN2; 59.2%). The three least frequently endorsed responses were: "I threaten the person who used hate speech with violence" (REV3; 5.7%), "I laugh at the statement" (REIN1; 6.5%), and "I cheer on the person making hate speech" (REIN2; 6.9%; see Supporting Information: Table S1). *Comforting the victim* ( $M = 3.55$ ,  $SD = 1.17$ ) was the most frequently endorsed subscale, followed by *countering hate speech* ( $M = 3.29$ ,  $SD = 1.05$ ), *ignoring* ( $M = 2.74$ ,  $SD = 1.14$ ), *seeking help at school* ( $M = 2.69$ ,  $SD = 1.18$ ), *helplessness* ( $M = 2.25$ ,  $SD = 0.98$ ), *revenge* ( $M = 1.81$ ,  $SD = 0.98$ ), and *reinforcing* ( $M = 1.65$ ,  $SD = 0.78$ ; see Table 4).

### 3.3 | Measurement invariance testing and differences as a function of demographics

The results of the measurement invariance testing with the full sample are reported in Table 3. Considering changes in goodness-of-

fit indices for grade, gender, and immigrant background, no substantial reduction (i.e.,  $\Delta CFI > 0.010$  and  $\Delta RMSEA > 0.015$ ) in model fit was found between the configural and metric invariance model and between the metric and scalar invariance model. These findings indicate scalar measurement invariance of the MBRHS across these groups and imply that latent means can be compared across those groups. Additional invariance testing (i.e., for SES and country of origin), which was not the main focus of the present study, is presented in the Supporting Information: Table S2.

#### 3.3.1 | Grade

The scalar measurement invariance model (Model 1, Table 3) was used to compare latent means across grades. Results showed that ninth graders showed lower means in *comforting the victim* than eighth graders ( $p < .001$ ,  $d = -0.16$ ) and seventh graders ( $p < .001$ ,  $d = -0.25$ ). Likewise, ninth graders reported lower means in *seeking help at school* compared with eighth graders ( $p < .001$ ,  $d = -0.26$ ) and seventh graders ( $p < .001$ ,  $d = -0.53$ ). In addition, eighth graders reported lower means of *seeking help at school* than seventh graders ( $p < .001$ ,  $d = -0.39$ ). Finally, ninth graders reported higher means in *reinforcing* compared with eighth graders ( $p = .003$ ,  $d = 0.11$ ).

#### 3.3.2 | Gender

The scalar measurement invariance model (Model 2, Table 3) was used to compare latent means between girls and boys. Results showed that, compared with boys, girls reported higher means in

**TABLE 3** Measurement Invariance Testing of the MBRHS across Grade, Gender, and Immigrant Background.

Models	$\chi^2$ (df)	$p$	RMSEA	$\Delta$ RMSEA	CFI	$\Delta$ CFI	Invariance rule accepted
<i>Model 1: Grade</i>							
Configural	1836.75 (504)	<.001	0.050		0.953		
Metric	1874.38 (530)	<.001	0.049	-0.001	0.952	-0.001	Yes
Scalar	1982.30 (558)	<.001	0.049	0.000	0.949	-0.003	Yes
<i>Model 2: Gender</i>							
Configural	1611.94 (336)	<.001	0.049		0.952		
Metric	1682.85 (349)	<.001	0.049	0.000	0.949	-0.003	Yes
Scalar	1884.17 (363)	<.001	0.052	0.003	0.943	-0.006	Yes
<i>Model 3: Immigrant background</i>							
Configural	1559.82 (336)	<.001	0.048		0.956		
Metric	1570.97 (349)	<.001	0.047	-0.001	0.956	0.000	Yes
Scalar	1632.89 (363)	<.001	0.047	0.000	0.954	-0.002	Yes

Abbreviations:  $\chi^2$ , chi-square test of model fit;  $df$ , degrees of freedom; CFI, comparative fit index;  $\Delta$ CFI, change in CFI compared with the weaker measurement invariance model above; RMSEA, root mean square error of approximation;  $\Delta$ RMSEA, change in RMSEA compared to the weaker measurement invariance model above.

comforting the victim ( $p < .001$ ,  $d = 0.68$ ), seeking help at school ( $p < .001$ ,  $d = 0.37$ ), and countering hate speech ( $p < .001$ ,  $d = 0.39$ ). Compared with boys, girls reported lower means in revenge ( $p < .001$ ,  $d = -0.54$ ), reinforcing ( $p < .001$ ,  $d = -0.52$ ), and ignoring ( $p < .001$ ,  $d = -0.17$ ).

### 3.3.3 | Immigrant background

The scalar measurement invariance model (Model 3, Table 3) was used to compare latent means between adolescents with and without an immigrant background. Results showed that, compared with adolescents without an immigrant background, adolescents with an immigrant background reported lower means in *comforting the victim* ( $p < .001$ ,  $d = -0.09$ ), *seeking help at school* ( $p < .001$ ,  $d = -0.14$ ), and *countering hate speech* ( $p < .001$ ,  $d = -0.09$ ). In comparison to those without an immigrant background, adolescents with an immigrant background reported higher means in *revenge* ( $p < .001$ ,  $d = 0.31$ ), *reinforcing* ( $p < .001$ ,  $d = 0.22$ ), and *ignoring* ( $p < .001$ ,  $d = 0.17$ ).

### 3.4 | Associations between empathy, moral disengagement, and bystander responses to racist hate speech

As shown in Table 4, *empathy toward victims of hate speech* was positively correlated with *comforting the victim* ( $r = .69$ ,  $p < .001$ ), *seeking support at school* ( $r = .51$ ,  $p < .001$ ), and *countering hate speech* ( $r = .57$ ,  $p < .001$ ), while being negatively correlated with *revenge* ( $r = -.18$ ,  $p < .001$ ), *reinforcing* ( $r = -.41$ ,  $p < .001$ ), *helplessness* ( $r = -.16$ ,  $p < .001$ ), and *ignoring* ( $r = -.22$ ,  $p < .001$ ). Conversely, *moral disengagement* was negatively correlated with *comforting the victim* ( $r = -.47$ ,  $p < .001$ ), *seeking support at school* ( $r = -.32$ ,  $p < .001$ ), and *countering hate speech* ( $r = -.40$ ,  $p < .001$ ), while being positively correlated with *revenge* ( $r = .35$ ,  $p < .001$ ), *reinforcing* ( $r = .64$ ,  $p < .001$ ), *helplessness* ( $r = .34$ ,  $p < .001$ ), and *ignoring* ( $r = .36$ ,  $p < .001$ ).

## 4 | DISCUSSION

Despite the growing number of research studies dealing with hate speech among adolescents, there is a lack of psychometrically sound instruments to measure bystander responses to hate speech. Hence, the present study aimed to shed light on the little understood role of bystanders in hate speech by (a) validating the scores from an instrument to measure bystander responses to racist hate speech among adolescents in schools, (b) testing measurement invariance and exploring latent mean differences in the various responses, and (c) investigating the relationships between empathy, moral disengagement, and bystander responses to hate speech.

### 4.1 | Psychometric properties, frequencies of bystander responses, and latent mean differences

The current study demonstrated that the MBRHS has good psychometric properties. EFA and CFA indicated that a model with seven factors could be confirmed. In addition to this, the analyses confirmed the validity and reliability of the MBRHS scores. These findings support the multidimensional nature of bystander response to hate speech among adolescents. That is, different reactions of bystanders can be identified and differentiated.

The subscale that received the highest rating was *comforting the victim*, followed by *countering hate speech*, *ignoring*, *seeking help at school*, *helplessness*, *revenge*, and *reinforcing*. These findings are comparable to research on coping with online hate speech victimization and bystander responses to bias-based bullying (Gámez Guadix et al., 2020; Gönültaş & Mulvey, 2021; Wachs et al., 2020). While the finding that two out of the three most frequently endorsed responses are prosocial is encouraging, this study also showed that *ignoring* was the third most reported response. The latter result, in combination with the fact that *seeking help in school* was seldom endorsed, presents a substantial challenge for schools responsible for the welfare of adolescents because such inaction can lead to personal

**TABLE 4** Descriptives and latent correlation matrix of the MBRHS, empathy, and moral disengagement.

Variable	M (SD)	1.	2.	3.	4.	5.	6.	7.	8.	
1. Comforting the victim	3.55 (1.17)	-								
2. Seeking help at school	2.69 (1.18)	0.56*	-							
3. Countering hate speech	3.29 (1.05)	0.61*	0.64*	-						
4. Revenge	1.81 (0.98)	-0.13*	-0.11*	-0.06**	-					
5. Reinforcing	1.65 (0.78)	-0.36*	-0.22*	-0.25*	0.64*	-				
6. Helplessness	2.25 (0.98)	-0.19*	-0.21*	-0.39*	0.03	0.26*	-			
7. Ignoring	2.74 (1.14)	-0.23*	-0.26*	-0.34*	0.08*	0.27*	0.48*	-		
8. Empathy	3.63 (0.96)	0.69*	0.51*	0.57*	-0.18*	-0.41*	-0.16*	-0.22*	-	
9. Moral disengagement	1.99 (0.77)	-0.47*	-0.32*	-0.40*	0.35*	0.64*	0.34*	0.36*	-0.56*	-

\* $p < .001$ ; \*\* $p < .05$ .



distress and health issues in bystanders (Bistrong et al., 2019). Hence, further research is needed to understand which barriers adolescents perceive as potentially explaining this kind of inactivity.

Testing multigroup measurement invariance of the MBRHS revealed strong evidence for scalar measurement invariance for grade, gender, and immigrant background, indicating an agreement on how bystander responses manifest between those groups. In addition to this, several differences were found in bystander responses as a function of demographic variables. Differences by grade were mainly small in magnitude but clearly showed that, compared with eighth and seventh graders, ninth graders reported lower means of *supporting the victim* and *seeking help at school*. Possible explanations for these differences might be that older adolescents are more likely to believe that victims of hate speech should be able to solve their problems independently and thus underestimate the negative consequences of hate speech. This idea is partially supported by research showing that older adolescents were more likely to play down the effects of race-based humor than younger adolescents (Mulvey et al., 2016). In addition, older students might be more reluctant to intervene, as they believe that even indirect interventions (e.g., *comforting the victim*) could make them a target and harm their social status among peers (Mulvey et al., 2016). Older adolescents' tendency to conform to specific hate speech-endorsing norms within the classroom might also explain why ninth graders reported reinforcing hate speech more frequently than eighth graders. Indeed, several studies have shown that peer pressure and social norms play a significant role in understanding classroom hate speech dynamics (Ballaschk et al., 2021; Wachs, Wettstein, et al., 2022). Overall, these findings are aligned with previous research on bullying that suggested that the action of defending a victim decreased with age and that of reinforcing the bully increased (Salmivalli & Voeten, 2004; Schultze-Krumbholz et al., 2018) but not with other research (e.g., Mazzone et al., 2016).

Gender differences were generally very moderate in magnitude and revealed a systematic pattern. More specifically, girls reported higher means for *comforting the victim*, *seeking help at school*, and *countering hate speech*, and lower means for *revenge and reinforcing* than boys. Girls also reported higher means than boys in *ignoring*. That is to say, girls showed a more positive response pattern to hate speech than boys. These findings also support traditional gender stereotypes in bystander behavior, which have been highlighted in related research on aggression and bullying, in which girls show more prosocial behavior than boys (Busching & Krahé, 2020; Mazzone et al., 2016; Yun & Graham, 2018), whereas boys show more pro-bully behavior than girls (Mazzone et al., 2016; Quenneville et al., 2022). This study suggests that gender-specific bystander behavior might also be found concerning hate speech.

Although the differences by immigrant background were generally very small in magnitude, the findings revealed that adolescents with immigrant background showed systematically higher values in challenging responses and lower means in socially desirable responses. More specifically, when compared with adolescents with immigrant background, those without immigrant

background reported higher levels of *comforting the victim*, *seeking support at school*, and *countering hate speech*. In addition, students without immigrant background reported lower levels of *revenge*, *reinforcing*, and *ignoring*. Young people with immigrant background often experience structural discrimination in German and Swiss schools (Gomolla, 2021). Such experiences can manifest in the belief that these students do not feel as if they are supported by teachers, which might explain why they are less likely to seek help, feel more helpless, and seek out revenge. The fear of becoming the next target when actively standing up against hate speech might also increase the likelihood of passive bystanding in students with an immigration background and the fact that they are less likely to report prosocial responses. The findings of the present study are inconsistent with research that suggests that adolescents without immigrant background showed lower odds of intervening in bias-based bullying when the target had an immigrant background (Gönültaş & Mulvey, 2021) and that adolescents without an immigrant background are more likely to intervene when the excluded peer also has no immigrant background (Palmer et al., 2022). An alternative explanation that needs further investigation is that the cultural context might explain these differences between previous research and our findings.

#### 4.2 | Associations between empathy, moral disengagement, and bystander responses

We found evidence that empathy is positively correlated with *comforting the victim*, *seeking help at school*, and *countering hate speech* but negatively associated with *revenge*, *reinforcing*, *helplessness*, and *ignoring*. The strongest correlation was found with *comforting the victim*, which seems reasonable as empathy may well be a prerequisite for this kind of response. The weakest correlation was found with *helplessness*, which might be explained by the fact that other factors (e.g., self-efficacy) might be more relevant in understanding bystanders' helplessness in response to hate speech. Overall, our findings are in line with research on (bias-based) traditional bullying and cyberbullying (e.g., Álvarez-García et al., 2021; Gönültaş & Mulvey, 2022; Schultze-Krumbholz et al., 2018).

Moral disengagement was negatively correlated with *comforting the victim*, *seeking help at school*, and *countering hate speech*, but positively correlated with *revenge*, *reinforcing*, *helplessness*, and *ignoring*. The strongest correlation was found between moral disengagement and *reinforcement*. This finding indicates that the deactivation of self-censoring mechanisms might play a particularly relevant role when adolescents witness hate speech and decide whether they reinforce hate speech—for example, by laughing about it or cheering on the person who uses it. The correlation coefficients with most other subscales were within the same range. However, there was a strong negative correlation between moral disengagement and *comforting the victim*, which indicates that moral disengagement might prevent bystanders from supporting victims of hate speech. These findings extend our knowledge in a similar

direction as that of research into the relationship between moral disengagement and (bias-based) traditional bullying and cyberbullying (Killer et al., 2019; Raboteg-Šaric, 2019; Shen et al., 2022; Thornberg et al., 2022).

### 4.3 | Limitations and future directions

Despite the new findings in this study, several limitations need to be acknowledged. First, some caution is advised when generalizing the results in this study to adolescents' responses to hate speech in general because the MBRHS focused on racist hate speech alone. More research is needed to distinguish bystander responses to different forms of hate speech (e.g., homophobic hate speech, antisemitic hate speech) by considering various correlates (being gay, being Jewish). Such research might help to understand the complexity of in-group and out-group processes when people decide whether or not to intervene in hate speech incidents, as well as to identify high-risk profiles. As well as considering different target groups separately, the intersection of other minority groups (e.g., being gay and having an immigrant background) should also be compared with the results of being part of just one minority group (e.g., being either gay or having an immigrant background).

Second, this study only included self-report measures and used a cross-sectional design. The exclusive use of self-reports might increase the risk of social desirability biases, which could be addressed with follow-up research based on peer nominations. The cross-sectional study design means that it is impossible to comprehend the temporal associations between empathy, moral disengagement, and the MBRHS. Follow-up longitudinal research could also help to test further aspects of validity (e.g., predictive validity, test-retest reliability) of the MBRHS.

Third, this study should be considered a first step in understanding adolescent bystander responses toward hate speech. Follow-up research should investigate more individual correlates (e.g., intergroup friendships, intergroup attitudes) and contextual correlates (e.g., inclusive classroom climate, classroom composition) that might influence adolescents' responses to hate speech.

Finally, the scores for the *reinforcement* subscale did not show good reliabilities. We thus recommend that this subscale needs to be further developed. To add to this, the MBRHS does not necessarily cover all potential bystander responses. Alongside *reinforcing*, one might consider including assisting the person who uses hate speech as an additional response in future research and which could thus represent another subscale. In its current form, the MBRHS considers the response *seeking help at school*. However, adolescents might also seek out support from parents or institutions outside of school. Therefore, we would also recommend that this might also be a valuable addition to the MBRHS.

### 4.4 | Practical implications

Despite these limitations, this study has important practical implications. First, one very encouraging finding was that adolescents often

used both indirect interventions (i.e., *comforting the victim*) and direct interventions (i.e., *countering hate speech*). These findings indicate that there is great potential for peer-to-peer approaches to educating adolescents in constructive techniques to deal with hate speech. Nonetheless, *ignoring* was also relatively common, suggesting a lack of skills in responding to hate speech or a lack of feeling responsible for intervening. Prevention programs should, therefore, aim to equip adolescents with the skills to stop them from ignoring incidents of hate speech. Addressing hate speech as a topic in schools might also increase adolescents' willingness to *seek help at school*.

Second, prevention programs for older adolescents should highlight the need to *comfort the victim* by raising awareness of the adverse outcomes of hate speech victimization. In addition, older adolescents should be informed about the possibilities of *seeking help at school*, with the benefits of having someone at school they could talk to being highlighted to them. At the same time, a particular focus in prevention programs should be on adolescents with immigrant background as well as boys because both provided responses that suggested unfavorable response patterns to hate speech.

Finally, prevention programs should include empathy and moral engagement training to increase prosocial and decrease antisocial or passive bystander responses. This concept is supported by research that showed that empathy training toward refugees was positively associated with a change in positive out-group attitudes and helpful behavior toward refugees among children (Taylor & Glen, 2020). Empathy training is also a part of the anti-hate speech prevention program "HateLess. Together against hatred," which effectively increases counterspeech directly and indirectly via empathy toward victims of hate speech (Wachs et al., 2023).

### ACKNOWLEDGMENTS

This work was supported by German Research Foundation (DFG) [grant numbers WA 4275/2-1 and BI 1046/9-1] and the University of Teacher Education Bern [grant number 19s 0008 01]. The research funding sources had no involvement in the conduct of the project or preparation of the present manuscript. Open Access funding enabled and organized by Projekt DEAL.

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

### DATA AVAILABILITY STATEMENT

The datasets analyzed during the current study are not publicly available but are available from the corresponding author upon reasonable request.

### ORCID

Sebastian Wachs  <http://orcid.org/0000-0003-2787-6646>

Ludwig Bilz  <http://orcid.org/0000-0003-3017-0492>

### REFERENCES

Álvarez-García, D., Thornberg, R., & Suárez-García, Z. (2021). Validation of a scale for assessing bystander responses in bullying. *Psicothema*, 33(4), 623–630. <https://doi.org/10.7334/psicothema2021.140>

- Ballaschk, C., Wachs, S., Krause, N., Schulze-Reichelt, F., Kansok-Dusche, J., Bilz, L., & Schubarth, W. (2021). Dann machen halt alle mit. "Eine qualitative studie zu beweggründen und motiven für hatespeech unter schüler\* innen. [Then everyone just goes along with it." A qualitative study on reasons and motives of hate speech among students]. *Diskurs Kindheits- und Jugendforschung*, 16(4), 13–14. <https://doi.org/10.3224/diskurs.v16i4.01>
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, 71(2), 364–374. <https://doi.org/10.1037/0022-3514.71.2.364>
- Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety, & W. Ickes (Eds.), *Social neuroscience. The social neuroscience of empathy* (pp. 3–15). MIT Press. <https://doi.org/10.7551/mitpress/9780262012973.003.0002>
- Bistrong, E., Bottiani, J. H., & Bradshaw, C. P. (2019). Youth reactions to bullying: Exploring the factors associated with students' willingness to intervene. *Journal of School Violence*, 18(4), 522–535. <https://doi.org/10.1080/15388220.2019.1576048>
- Boag, E. M., & Carnelley, K. B. (2016). Attachment and prejudice: The mediating role of empathy. *British Journal of Social Psychology*, 55(2), 337–356. <https://doi.org/10.1111/bjso.12132>
- Bundesministerium des Innern und Heimat (BMI) (2023). *Verfassungsschutzbericht 2022*. Bundesministerium des Innern und für Heimat. [https://www.verfassungsschutz.de/SharedDocs/publikationen/DE/verfassungsschutzberichte/2023-06-20-verfassungsschutzbericht-2022.pdf?\\_\\_blob=publicationFile&v=5](https://www.verfassungsschutz.de/SharedDocs/publikationen/DE/verfassungsschutzberichte/2023-06-20-verfassungsschutzbericht-2022.pdf?__blob=publicationFile&v=5)
- Busching, R., & Krahé, B. (2020). With a little help from their peers: The impact of classmates on adolescents' development of prosocial behavior. *Journal of Youth and Adolescence*, 49(9), 1849–1863. <https://doi.org/10.1007/s10964-020-01260-8>
- Castellanos, M., Wettstein, A., Wachs, S., & Bilz, L. (2023). Direct and indirect effects of social dominance orientation on hate speech perpetration via empathy and moral disengagement among adolescents: A multilevel mediation model. *Aggressive Behavior*. <https://doi.org/10.1002/ab.22100>
- Castellanos, M., Wettstein, A., Wachs, S., Kansok-Dusche, J., Ballaschk, C., Krause, N., & Bilz, L. (2023). Hate speech in adolescents: A binational study on prevalence and demographic differences. *Frontiers in Education*, 8, 1076249. <https://doi.org/10.3389/educ.2023.1076249>
- Celuch, M., Oksanen, A., Räsänen, P., Costello, M., Blaya, C., Zych, I., Llorent, V. J., Reichelmann, A., & Hawdon, J. (2022). Factors associated with online hate acceptance: A cross-national six-country study among young adults. *International Journal of Environmental Research and Public Health*, 19(1), 534. <https://doi.org/10.3390/ijerph19010534>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Clark, D. A., & Bowles, R. P. (2018). Model fit and item factor analysis: Overfactoring, underfactoring, and a program to guide interpretation. *Multivariate Behavioral Research*, 53(4), 544–558. <https://doi.org/10.1080/00273171.2018.1461058>
- Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Inc.
- Cowan, G., & Khatchadourian, D. (2003). Empathy, ways of knowing, and interdependence as mediators of gender differences in attitudes toward hate speech and freedom of speech. *Psychology of Women Quarterly*, 27(4), 300–308. <https://doi.org/10.1111/1471-6402.00110>
- D'Errico, F., & Paciello, M. (2018). Online moral disengagement and hostile emotions in discussions on hosting immigrants. *Internet Research*, 28(5), 1313–1335. <https://doi.org/10.1108/IntR-03-2017-0119>
- Eisinga, R., Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, cronbach, or Spearman-Brown. *International Journal of Public Health*, 58(4), 637–642. <https://doi.org/10.1007/s00038-012-0416-3>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Finch, W. H. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis. *Educational and Psychological Measurement*, 80(2), 217–241. <https://doi.org/10.1177/0013164419865769>
- Fischer, S. M., John, N., Melzer, W., Kaman, A., Winter, K., & Bilz, L. (2020). Traditional bullying and cyberbullying among children and adolescents in Germany—Cross-sectional results of the 2017/18 HBSC study and trends. *Journal of Health Monitoring*, 5(3), 53–68. <https://doi.org/10.25646/6902>
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science/Revue Canadienne des sciences du comportement*, 49(2), 78–88. <https://doi.org/10.1037/cbs000069>
- Gámez Guadix, M., Wachs, S., & Wright, M. (2020). "Haters back off!" psychometric properties of the coping with cyberhate questionnaire and relationship with well-being in spanish adolescents. *Psicothema*, 32(4), 567–574. <https://doi.org/10.7334/psicothema2020.219>
- Gomolla, M. (2021). Educational inequalities, vulnerability for discrimination and the politics of school change in Germany as a post-migration society. In N. Fromm, A. Jünemann, & H. Safouane (Eds.), *Power in vulnerability: A multi-dimensional review of migrants' vulnerabilities* (pp. 175–198). Springer VS. [https://doi.org/10.1007/978-3-658-34052-0\\_9](https://doi.org/10.1007/978-3-658-34052-0_9)
- Gönültaş, S., & Mulvey, K. L. (2021). The role of immigration background, intergroup processes, and social-cognitive skills in bystanders' responses to bias-based bullying toward immigrants during adolescence. *Child Development*, 92(3), 296. <https://doi.org/10.1111/cdev.13476>
- Gönültaş, S., & Mulvey, K. L. (2022). Do adolescents intervene in intergroup bias-based bullying? Bystander judgments and responses to intergroup bias-based bullying of refugees. *Journal of Research on Adolescence*, 33(1), 4–23. <https://doi.org/10.1111/jora.12752>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Pearson College Division.
- Hartley, J. E. K., Levin, K., & Currie, C. (2016). A new version of the HBSC family affluence scale-FAS III: Scottish qualitative findings from the international FAS development study. *Child Indicators Research*, 9, 233–245. <https://doi.org/10.1007/s12187-015-9325-3>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kansok-Dusche, J., Ballaschk, C., Krause, N., Zeiðig, A., Seemann-Herz, L., Wachs, S., & Bilz, L. (2022). A systematic review on hate speech among children and adolescents: Definitions, prevalence, and overlap with related phenomena. *Trauma, Violence & Abuse*, 15248380221108070. <https://doi.org/10.1177/15248380221108070>
- Killer, B., Bussey, K., Hawes, D. J., & Hunt, C. (2019). A meta-analysis of the relationship between moral disengagement and bullying roles in youth. *Aggressive Behavior*, 45(4), 450–462. <https://doi.org/10.1002/ab.21833>
- Knauf, R.-K., Eschenbeck, H., & Hock, M. (2018). Bystanders of bullying: Social-cognitive and affective reactions to school bullying and cyberbullying. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(4), Article 3. <https://doi.org/10.5817/CP2018-4-3>

- Krause, N., Ballaschk, C., Schulze-Reichelt, F., Kansok-Dusche, J., Wachs, S., Schubarth, W., & Bilz, L. (2021). „Ich lass mich da nicht klein machen!“ eine qualitative studie zur bewältigung von hatespeech durch schüler/innen. [“I Don't let them get me down!”—A qualitative study on students' coping with hate speech]. *Zeitschrift für Bildungsforschung*, 11(1), 169–185. <https://doi.org/10.1007/s35834-021-00291-w>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Mazzone, A., Camodeca, M., & Salmivalli, C. (2016). Interactive effects of guilt and moral disengagement on bullying, defending and outsider behavior. *Journal of Moral Education*, 45(4), 419–432. <https://doi.org/10.1080/03057240.2016.1216399>
- Mulvey, K. L., Palmer, S. B., & Abrams, D. (2016). Race-based humor and peer group dynamics in adolescence: Bystander intervention and social exclusion. *Child Development*, 87(5), 1379–1391. <https://doi.org/10.1111/cdev.12600>
- Muthén, L. K., & Muthén, B. O. (2012–2021). *Mplus User's Guide*. Muthén & Muthén.
- Nocentini, A., Menesini, E., & Salmivalli, C. (2013). Level and change of bullying behavior during high school: A multilevel growth curve analysis. *Journal of Adolescence*, 36(3), 495–505. <https://doi.org/10.1016/j.adolescence.2013.02.004>
- Palmer, S. B., Filippou, A., Argyri, E. K., & Rutland, A. (2022). Minority-and majority-status bystander reactions to, and reasoning about, intergroup social exclusion. *Journal of Experimental Child Psychology*, 214, 105290. <https://doi.org/10.1016/j.jecp.2021.105290>
- Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, 38(6), 922–934. <https://doi.org/10.1002/ejsp.504>
- Quenneville, S., Talwar, V., & Bosacki, S. (2022). Teacher ratings and adolescent students' perceived social behaviours and gender-role orientations. *Journal of Gender Studies*, 31(4), 444–456. <https://doi.org/10.1080/09589236.2021.1988530>
- Raboteg-Šaric, Z. (2019). Empathy and moral disengagement as predictors of bystander roles in school bullying. *The Central European Journal of Paediatrics*, 15(2), 161–176. <https://doi.org/10.5457/p2005-114.248>
- Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior*, 22(1), 1–15. [https://doi.org/10.1002/\(SICI\)1098-2337\(1996\)22:1<1::AID-AB1>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2337(1996)22:1<1::AID-AB1>3.0.CO;2-T)
- Salmivalli, C., & Voeten, M. (2004). Connections between attitudes, group norms, and behaviour in bullying situations. *International Journal of Behavioral Development*, 28(3), 246–258. <https://doi.org/10.1080/01650250344000488>
- Schultze-Krumbholz, A., Hess, M., Pfetsch, J., & Scheithauer, H. (2018). Who is involved in cyberbullying? Latent class analysis of cyberbullying roles and their associations with aggression, self-esteem, and empathy. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(4), 2. <https://doi.org/10.5817/CP2018-4-2>
- Shen, Y., Yuan, L., Xiong, X., & Xin, T. (2022). Empathy and cyberbystander behavior: The role of moral disengagement. *Current Psychology*, 1–10. <https://doi.org/10.1007/s12144-022-03015-z>
- Taylor, L. K., & Glen, C. (2020). From empathy to action: Can enhancing host-society children's empathy promote positive attitudes and prosocial behaviour toward refugees? *Journal of Community & Applied Social Psychology*, 30(2), 214–226. <https://doi.org/10.1002/casp.2438>
- Teerenstra, S., Lu, B., Preisser, J. S., van Achterberg, T., & Borm, G. F. (2010). Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics*, 66(4), 1230–1237. <https://doi.org/10.1111/j.1541-0420.2009.01374.x>
- Thornberg, R., Pozzoli, T., & Gini, G. (2022). Defending or remaining passive as a bystander of school bullying in Sweden: The role of moral disengagement and antibullying class norms. *Journal of Interpersonal Violence*, 37(19–20), 18666. <https://doi.org/10.1177/08862605211037427>
- Wachs, S., Bilz, L., Wettstein, A., Wright, M. F., Kansok-Dusche, J., Krause, N., & Ballaschk, C. (2022). Associations between witnessing and perpetrating online hate speech among adolescents: Testing moderation effects of moral disengagement and empathy. *Psychology of Violence*, 12(6), 371–381. <https://doi.org/10.1037/vio0000422>
- Wachs, S., Gámez-Guadix, M., Wright, M. F., Görzig, A., & Schubarth, W. (2020). How do adolescents cope with cyberhate? Psychometric properties and socio-demographic differences of a coping with cyberhate scale. *Computers in Human Behavior*, 104, 106167. <https://doi.org/10.1016/j.chb.2019.106167>
- Wachs, S., Krause, N., Wright, M. F., & Gámez-Guadix, M. (2023). Effects of the prevention program “HateLess. together against hatred” on adolescents' empathy, self-efficacy, and countering hate speech. *Journal of Youth and Adolescence*, 52(6), 1115–1128. <https://doi.org/10.1007/s10964-023-01753-2>
- Wachs, S., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2022). Adolescents' motivations to perpetrate hate speech and links with social norms. *Comunicar*, 30(71), 9–20. <https://doi.org/10.3916/C71-2022-01>
- Yang, S. A., & Kim, D. H. (2017). Factors associated with bystander behaviors of Korean youth in school bullying situations: A cross-sectional study. *Medicine*, 96(32), e7757. <https://doi.org/10.1097/MD.0000000000007757>
- Yates, F., & Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society*, 15(2), 253–261. <https://www.jstor.org/stable/2983772>
- Yun, H. Y., & Graham, S. (2018). Defending victims of bullying in early adolescence: A multilevel analysis. *Journal of Youth and Adolescence*, 47(9), 1926–1937. <https://doi.org/10.1007/s10964-018-0869-7>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wachs, S., Bilz, L., Wettstein, A., & Espelage, D. L. (2023). Validation of the multidimensional bystander responses to racist hate speech scale and its association with empathy and moral disengagement among adolescents. *Aggressive Behavior*, 1–12. <https://doi.org/10.1002/ab.22105>